AI-Generated Feedback in English Writing Instruction for Language Learners: A Systematic Review

Inanc Karagoz Bartin University

ABSTRACT

This systematic review examines 22 studies (2024–2025) on the use of generative AI, primarily ChatGPT, for providing feedback in English writing instruction for language learners. It identifies the types of feedback AI offers, its effectiveness relative to teacher and peer feedback, and perceptions from students and teachers. Findings show AI excels in addressing surface-level issues (grammar, vocabulary, structure) and is valued for speed and comprehensiveness. However, both students and teachers emphasize the continued importance of human feedback for deeper, contextual revisions. While educators recognize AI's potential to ease workloads and improve feedback timeliness, concerns about reliability and depth persist. The review recommends integrating AI feedback with teacher oversight, training in prompt literacy and feedback evaluation, and calls for further research on long-term impacts, scaffolding strategies, and ethical considerations.

INTRODUCTION

Generative AI tools like ChatGPT can facilitate adaptive and personalized learning environments by compiling educational content compatible with learners' varying needs (Abbas et al., 2023). Additionally, real-time feedback and interactivity can improve understanding. Such flexibility and interactivity constitute a significant supplement to traditional practices (Abbas et al., 2023). When it comes to the instruction of writing skills, automated writing evaluation (AWE) has been a prominent research vein well before the availability of large language models. This review excludes such research as it deems the dialogic, iterative, and creative output (Maphoto et al., 2024; Wang, 2024) of generative AI tools a distinctive trait, as opposed to grammar and spelling check with scores solely based on what is fed to them. On examining how these two modalities perform in educational settings, generative AI and AWE both focus on improving writing quality but utilize fundamentally different approaches. Traditional automated writing instruction often emphasizes predefined metrics and static feedback for assessments (Rahman et al., 2022). In contrast, generative AI makes use of vast databases of linguistic input to generate contextually relevant text, allowing students to engage with writing as a more iterative and constructive process (Maphoto et al., 2024; Wang, 2024). This shift towards an immersive interactive experience nurtures new pathways for students' engagement with writing tasks, cultivating their writing skills in a more organic manner, distinct from the assessment-centric nature of former automated writing tools. Generative AI's interactive layer can make meaning negotiation and consideration of different perspectives possible. Learners' ability to ask clarifying questions and receive explanations has potential for filling gaps in mastery of content, organization, and mechanics. It has been reported that students possessed a favorable attitude toward the integration of generative AI in learning, thanks to personalized support in writing and brainstorming tasks (Chan & Hu, 2023).

This potential for generative AI to transform writing instruction extends beyond its capacity to provide personalized feedback—it also addresses a persistent challenge faced by educators. Delivering constructive, timely feedback on student writing is inherently complex, as teachers must consider both the content and process aspects of writing, while also deciding the most effective source of feedback. By incorporating AI-generated feedback into the classroom, teachers can alleviate some of this burden, allowing them to focus on facilitating higher-order critical thinking and individualized instruction. In this light, generative AI tools serve as an invaluable complement to traditional teaching methods, offering rapid and context-sensitive feedback that can spark further revisions and dialogue in the classroom.

Ethical concerns and risks of plagiarism are often addressed in educational research (Simms, 2024; Yeralan & Lee, 2023). Students' growing reliance on work generated by AI tools calls for reconsideration of how educators assess student learning. Since their use has permeated different aspects of academic and daily life, teaching students how to distinguish credible information and filter it through critical thinking emerges as an important focus of AI-supported learning (Simms, 2024). Previously, Teng (2024) conducted a systematic review of the use of ChatGPT in EFL writing instruction. He framed ChatGPT as a revolutionary tool that not only supports writing but also necessitates the development of AI literacy and digitalized writing skills. In contrast, the current review narrows this scope, focusing exclusively on AI-generated feedback. It investigates how students and teachers perceive the feedback provided by such tools, aiming to elucidate its practical effectiveness in enhancing the revision process and overall writing performance.

This paper is organized into four sections. Following this introduction, the Methodology outlines the systematic review approach, including the search strategy, screening, and selection process, in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). The Findings and Discussion section analyzes key themes, trends, and challenges in the use of AI-generated feedback. Finally, the Conclusion discusses the broader implications for educational practice and future research, offering recommendations for the effective integration of AI tools while preserving pedagogical integrity.

METHODOLOGY

This section presents the research questions, inclusion and exclusion criteria, search strategies, and the coding process. The following research questions guide this review:

- 1. What types of feedback do generative AI tools provide on students' writing in English learning contexts?
- 2. How does AI-generated feedback compare to teacher or peer feedback in these contexts?
- 3. What are the benefits and challenges of employing AI-generated feedback in English writing instruction?
- 4. What are English language learners' and teachers' attitudes toward using generative AI tools for feedback on their writing?

Inclusion and Exclusion Criteria

Studies incorporating AI-generated feedback in second language writing from all age groups are included in this review. Experimental, exploratory, and descriptive studies are considered, provided they are published in English in a scholarly journal. Conference papers and review articles are excluded. Studies that broadly examine generative AI in EFL or ESL contexts, without a specific focus on AI-generated feedback, are also excluded. A chronological filter is not applied, as generative AI is an emerging technology, and its implementation in L2 settings remains limited. In short, studies are included in the final pool if they meet the following criteria:

- 1. The study examined the use of a generative AI tool (e.g. ChatGPT, Bard/Gemini, CoPilot) in writing instruction.
- 2. The study was conducted in an ESL or an EFL context.
- 3. The study focused on feedback on students' writing.
- 4. Participants were English language learners.

Search Strategies

In accordance with the PRISMA guidelines (Page et al., 2021), a comprehensive literature search was conducted to identify empirical studies examining the role of generative AI in providing writing feedback in English as a Foreign Language (EFL) contexts. Two academic databases were scanned to extract relevant studies: Scopus and Web of Science. These databases cover a comprehensive range of peer-reviewed articles in the field of education. A combination of keywords and Boolean operators was used in advanced search mode to capture relevant studies. The terms *generative AI*, *English language learning*, *writing*, and *feedback* were the main pillars of the search, along with their alternative wordings. The search query for each database was as follows:

Table 1

Query Used on Web of Science

(TI=("generative AI" OR "artificial intelligence" OR "AI-assisted writing" OR "AI writing tools" OR "ChatGPT" OR "GPT-4" OR "AI" OR "Gemini" OR "Bard" OR "Copilot") AND TS=("English as a Foreign Language" OR "EFL" OR "English language learners" OR "ESL" OR "second language") AND TI=("writing instruction" OR "academic writing" OR "writing skills" OR "writing") AND TI=("feedback"))

Table 2

Query Used on Scopus

(TITLE ("generative AI" OR "artificial intelligence" OR "AI-assisted writing" OR "AI writing tools" OR "ChatGPT" OR "GPT-4" OR "AI" OR "Gemini" OR "Bard" OR "Copilot") AND ABS ("English as a Foreign Language" OR "EFL" OR "English language learners" OR "ESL" OR "second language") AND ABS ("writing instruction" OR "academic writing" OR "writing skills") AND TITLE-ABS-KEY (feedback))

The search query led to 19 results in Web of Science, and 32 results in Scopus. When exclusion criteria were applied and five duplicates were removed, a total of 46 articles remained, with their abstracts retrieved. Studies were considered eligible if they focused on the use of generative AI tools for providing feedback in writing instruction for EFL learners and presented original empirical data. Records that focused solely on traditional automated writing evaluation or did not include empirical findings were excluded, resulting in the removal of 17 articles. The full texts of the remaining 29 studies were retrieved and evaluated for eligibility. Studies were excluded at this stage if they failed to adequately address the research questions of this review or if methodological limitations compromised their relevance. This phase resulted in the exclusion of seven additional studies, with a final selection of 22 studies (Figure 1).



Figure 1. PRISMA Flow Diagram Based on Page et al. (2021).

Coding

In addition to including the identifying features such as author(s)' name(s), publication year, and title, the following features were coded while reviewing the final pool of studies: 'Research questions' outline the primary issues and objectives addressed by each study. 'Setting' records basic information about the context of the study, including the age group of participants and the location of the instruction. The 'proficiency' level of participants (e.g., intermediate, advanced) is noted to highlight potential differences in how AI-generated feedback operates across various learner levels. 'Participants' refers to the sample size and demographic information, indicating the

breadth and characteristics of the study population. Details regarding the nature and duration of the 'intervention' are captured to understand how AI tools were integrated into the writing instruction. 'Feedback types' categorizes the forms of feedback provided by AI tools (e.g., surface-level, content, grammar, corrective, directive, informative). The code 'data sources' records the types of data (e.g., questionnaires, writing samples, system logs) collected. 'Data analysis' lists the methods employed to analyze the data. 'Key findings' summarize the main outcomes of the study, and 'implications' identify the broader educational and practical consequences derived from the findings. Initial coding labels (Table 3) were helpful in analyzing the studies in their entirety, as they provided comprehensive data about the interventions, key variables, and methods.

Table 3

Initial Coding Scheme

Author(s)	Year	Title
Research Questions	Setting	Proficiency
Participants	Intervention	Feedback Types
Data Sources Implications	Data Analysis	Key Findings

In order to summarize the key elements of the studies selected for this review, the following seven labels in the initial coding scheme were removed: title, research questions, proficiency, procedure, feedback types, data sources, and data analysis. The codes author(s) and year were merged in this iteration (Table 4). The remaining codes were retained in the summary.

Table 4

Codes Selected for the Summary

Author(s) & Year	Setting	Participants
Intervention	Key Findings	

In light of the research questions of this review, the findings of the selected studies are analyzed across four key areas: (1) the types of feedback provided by generative AI tools in English writing contexts; (2) the effectiveness of AI-generated feedback on writing performance; (3) learners' and instructors' perceptions of the utility and limitations of AI-generated feedback; and (4) the ways in which AI-generated feedback is integrated into instructional practices. The publication date was not used as a criterion for inclusion or exclusion, and all studies in the final pool were published from 2024 onward.

FINDINGS AND DISCUSSION

Following the summary provided in Table 5, this section presents the findings in five subsections: an overview of the selected studies, the role of generative AI feedback in EFL writing tasks, the impact of AI-generated feedback on writing performance, students' and teachers' perceptions of AI-generated feedback, and the challenges associated with AI-generated feedback.

Authors	Setting	Participants	Intervention	Key Findings
Abduljawad (2024)	Saudi Arabia, University	130 students	1-month AI-integrated writing module focusing on grammar, structure, and vocabulary	Significant writing improvement. AI useful but impersonal; concerns over creativity loss.
Allen & Mizumoto (2024)	Japan, University	33 students	4-week study comparing ChatGPT and peer review for proofreading	Students preferred ChatGPT for its clear, specific, and easily implementable feedback over peer review.
Alsofyani & Barzanji (2024)	Saudi Arabia, University	102 students	8-week comparing ChatGPT and teacher feedback	Both teacher and ChatGPT feedback improved writing. Teacher feedback is valued more.
Asadi et al. (2025)	Iran, IELTS course	68 students, 8 teachers	12-session: ChatGPT + teacher feedback vs. teacher-only feedback	Integrated AI-teacher feedback significantly outperformed teacher-only feedback across all IELTS criteria.
Bacon & Maneerutt (2024)	Thailand, University	143 students	15-week writing course integrating ChatGPT feedback & peer-assisted learning (PAL)	AI feedback was immediate and useful. PAL enhanced confidence and engagement. Writing improved in grammar and organization.
Ghafouri et al. (2024)	Iran, Online Course	12 teachers, 48 students	10-week ChatGPT-based Writing Instruction Protocol	AI supported protocol produced higher test scores and boosted teachers' self efficacy.
Guo & Wang (2024)	China, University	5 teachers	Comparison of ChatGPT and teacher feedback on 50 argumentative essays	ChatGPT feedback was more directive and praise-oriented. Teachers' feedback was more informative and inquiry-based.
Guo et al. (2024)	China, University	124 students	9 peer review tasks over 3 weeks; AI-supported vs. traditional peer review	AI-assisted peer reviewers gave better quality feedback.
Hwang et al. (2024)	Korea, University	11 students	3-week study on learner prompting behavior in ChatGPT-assisted writing revision	Students improved grammar but struggled with higher-order revisions. They use generic prompts.
Jamshed et al. (2024)	India, High School	132 students	8-week study comparing ChatGPT mobile feedback vs. teacher feedback	AI feedback by the ChatGPT mobile app led to better grammar accuracy.

Table 5Summary of the Selected Articles

Kurt & Kurt (2024)	Türkiye, University	52 preservice teachers	15-week study comparing ChatGPT, peer, and teacher feedback	ChatGPT delivered detailed feedback, though its quality varied with prompt quality.
Mahapatra (2024)	India, University	72 students	6-hour ChatGPT feedback intervention for self/peer assessment	AI feedback group showed significant immediate and delayed improvements in content and grammar.
Mohammed & Khalid (2025)	Iraq, University	322 students	10-week online writing course; ChatGPT vs. teacher feedback	AI boosted motivation, reduced anxiety; better writing and emotional intelligence.
Polakova & Ivenz (2024)	Czech Republic, University	110 students	Quasi-experimental study on ChatGPT vs. no-AI feedback over a semester	Grammar, conciseness, and inclusion of key information improved; AI feedback had some inaccuracies.
Teng (2025)	China, University	40 students	15-week writing course comparing ChatGPT vs. traditional feedback	Higher motivation and engagement with AI. Strong metacognitive awareness linked to effective AI use.
Tseng & Lin (2024)	Taiwan, University	15 students	ADDIE-based course integrating ChatGPT for writing feedback	Improvement in organization and coherence. AI accelerated revision and substituted traditional peer reviewers.
Wale & Kassahun (2024)	Ethiopia, University	92 students	AI-integrated writing tasks using Writerly and Google Docs	AI-integrated tasks led to significant improvements in grammar and organization.
Werdiningsi h et al. (2024)	Indonesia, University	3 students	ChatGPT integrated into academic writing process	AI useful for structure and vocabulary. Concerns over loss of personal voice.
Yan (2024)	China, University	117 students,6 teachers	7-week: individual vs. collaborative processing of ChatGPT feedback	Individual AI feedback improved short-term writing. Teacher scaffolding was key for long-term development.
Yao et al. (2025)	China, High School	13 teachers	2-week AI chatbot vs. teacher feedback study	AI feedback excelled in grammar corrections. Teacher feedback remained superior for deep, contextual analysis.
Zeevy- Soloyev (2024)	Israel, University	30 students	Comparison of peer, ChatGPT, and teacher writing feedback	ChatGPT feedback was clear, but teacher feedback was ultimately preferred for personalized explanations.
Zou et al. (2025)	China, University	20 students	Comparison of ChatGPT and teacher feedback uptake	Teacher feedback was adopted and trusted more; ChatGPT suggestions were less nuanced but improved organization.

An Overview of Selected Studies

Given the novelty of generative AI use in education, studies examining its application for writing feedback have only recently been published. Seventeen studies appeared in 2024 and five in 2025. These studies span a diverse geographic range, covering 14 countries across multiple continents. Six studies were conducted in China; two each in Iran, India, and Saudi Arabia; and one each in the Czech Republic, Ethiopia, Indonesia, Iraq, Israel, Japan, Korea, Taiwan, Thailand, and Turkiye (Figure 2). Seventeen studies took place in university-level EFL courses, two in high schools, one in an online course, and another in a language institution (Figure 3).



Figure 2. Geographical Distribution of Studies.



Figure 3. Setting of Studies.

Regarding participant size, four categories were established to group the studies: 1-5, 6-30, 31-70, and 71+ participants (Figure 4). Notably, the largest group comprises studies with 71 or more participants (11 studies), indicating that many researchers opted for larger samples to enhance the statistical robustness and generalizability of their findings. In contrast, only two studies involved very small sample sizes (1-5 participants), typically reflecting in-depth qualitative investigations. The mid-range categories—6-30 and 31-70 participants, which include five and four studies, respectively—were characteristic of quantitative and mixed-method designs. This distribution highlights a general trend toward larger sample sizes while still accommodating smaller, more exploratory studies.



Figure 4. Sample size distribution of the studies.

Most studies adopted a mixed-methods research design to capture both quantifiable patterns and in-depth insights for effective feedback. Thirteen studies followed this approach, five employed qualitative methodology, and four adopted a quantitative design (Table 6). Although the reviewed studies predominantly utilized mixed-methods approaches with sizable participant samples, a clear methodological pattern emerges: studies commonly relied on short-term interventions, limiting the generalizability of their findings. Intervention durations ranged from brief workshops (Mahapatra's 6-hour session, n = 72) to semester-long courses (Teng's 15-week study, n = 40), complicating cross-study comparisons.

Nineteen studies employed ChatGPT, while the alternative systems Kimi, Eva, and Writerly were each used in only one study (Figure 5). Although Copilot and Gemini were included in the search terms, no results were retrieved. This indicates that ChatGPT is the predominant AI tool in this area, suggesting its broader accessibility or acceptance among researchers. In contrast, the relatively few studies utilizing other AI systems highlight emerging avenues for future research, where comparative evaluations could determine whether these alternatives offer distinct advantages in providing writing feedback.

Table 7 presents the studies in terms of the recruited participants. Most studies recruited only students, indicating a predominant focus on student experiences with AI-generated feedback. The scarcity of studies involving teachers exclusively suggests that students' perceptions and experiences remain central to the alignment of generative AI tools with learner needs. Additionally, a subset of studies included both teachers and students, offering a more comprehensive perspective of the feedback process from multiple stakeholder viewpoints.

Table 6Methodologies of Selected Studies

Methodology	Studies
Quantitative	Allen & Mizumoto (2024); Guo et al. (2024); Jamshed et al. (2024); Yan (2024)
Qualitative	Kurt & Kurt (2024); Tseng & Lin (2024); Werdiningsih et al. (2024); Yao et al. (2025); Zeevy-Soloyev (2024)
Mixed-method	Hwang et al. (2024); Ghafouri et al. (2024); Mahapatra (2024); Abduljawad (2024); Alsofyani & Barzanji (2024); Polakova & Ivenz (2024); Guo & Wang (2024); Zou et al. (2025); Teng (2025); Asadi et al. (2025); Mohammed & Khalid (2025); Wale & Kassahun (2024); Bacon & Maneerutt (2024)

Table 7

Participants of Selected Studies

Participants	Studies
Students Only	Allen & Mizumoto (2024); Zeevy-Soloyev (2024); Hwang et al. (2024); Werdiningsih et al. (2024); Tseng & Lin (2024); Mahapatra (2024); Abduljawad (2024); Alsofyani & Barzanji (2024); Polakova & Ivenz (2024); Zou et al. (2025); Teng (2025); Mohammed & Khalid (2025); Guo et al. (2024); Wale & Kassahun (2024); Bacon & Maneerutt (2024); Jamshed et al. (2024)
Pre-Service EFL Teachers	Kurt & Kurt (2024)
Teachers Only	Yao et al. (2025); Guo & Wang (2024)
Both Students and Teachers	Ghafouri et al. (2024) [12 Teachers, 48 Students]; Yan (2024) [6 Teachers, 117 Students]; Asadi et al. (2025) [8 Teachers, 68 Students]



Figure 5. Distribution of Studies by AI System

Situating AI-Generated Feedback in EFL Writing Tasks

The primary research foci in the reviewed articles are presented in Table 8. Most studies focus on examining the process and outcomes of incorporating AI-generated feedback into EFL writing instruction. Teng (2025) further explored metacognitive awareness to gain a deeper understanding of the effective use of AI-generated feedback.

Table 8

teachers

assistance

Teacher–AI integration protocols

0 0	
Foci of Research	Studies
Corrective feedback for editing & proofreading	Allen & Mizumoto (2024); Jamshed et al. (2024); Mohammed & Khalid (2025); Zeevy-Soloyev (2024); Werdiningsih et al. (2024)
Prompts created by learners	Hwang et al. (2024)
Both mechanical and organizational aspects of writing	Abduljawad (2024); Alsofyani & Barzanji (2024); Kurt & Kurt (2024); Mahapatra (2024); Polakova & Ivenz (2024); Tseng & Lin (2024); Yan (2024)
Taxonomy & classification of feedback	Guo & Wang (2024)
Comparative/integrated feedback: ChatGPT vs. teacher/peer feedback	Allen & Mizumoto (2024); Alsofyani & Barzanji (2024); Asadi et al. (2025); Jamshed et al. (2024); Mohammed & Khalid (2025); Solovey (2024); Yao et al. (2025); Zou et al. (2025)
Metacognitive & engagement effects of ChatGPT feedback	Teng (2025)
AI-guided chatbot feedback by	Yao et al. (2025)

Integration of AI-generated Feedback in Research Design

AI-supported peer feedback via chatbot Guo et al. (2024)

AI-generated feedback is balanced across content, organization, and mechanics, but it may also yield off-target suggestions (Alsofyani & Barzanji, 2024; Abduljawad, 2024; Guo & Wang, 2024; Kurt & Kurt, 2024, Zeevy-Soloyev, 2024). In contrast, teacher feedback is tailored to specific learning goals with references to nuances in both surface-level and deeper structural aspects of writing (Zeevy-Soloyev, 2024; Zou et al., 2025; Yao et al., 2025; Polakova & Ivenz, 2024). Such feedback is grounded in the teacher's understanding of students' progress and writing style (Zeevy-Soloyev, 2024; Asadi et al., 2025; Yao et al., 2025), but it requires thoughtful review and ample time, causing delays in feedback turnaround (Yao et al., 2025). AI feedback can address this issue through immediate feedback, leading to quick revisions (Abduljawad, 2024; Kurt & Kurt, 2024; Mohammed & Khalid, 2025; Jamshed et al., 2024). However, teacher oversight is emphasized (Yao et al., 2025) to avoid solely generic feedback that misses unique contextual

Asadi et al. (2025); Ghafouri, & Zafarghandi (2024)

factors (Alsofyani & Barzanji, 2024; Abduljawad, 2024; Guo & Wang, 2024; Zou et al., 2025). Although studies frequently illustrate AI's capacity for immediate, detailed corrections, the reviewed research collectively highlights a critical shortcoming: limited contextual sensitivity. The recurrent finding that human-generated feedback is helpful in addressing deeper content issues (e.g., Zou et al., 2025; Yao et al., 2025) points to a fundamental limitation of current generative AI tools, emphasizing the necessity of maintaining teacher oversight in feedback processes.

The Impact of AI-Generated Feedback on Writing Performance

This review shows that AI-generated feedback improved surface-level aspects of writing, such as vocabulary, grammar, and spelling (Abduljawad, 2024; Hwang et al., 2024; Jamshed et al., 2024). Improvements were also reported in terms of organization (Guo & Wang, 2024; Tseng & Lin, 2024; Zou et al., 2025). Abduljawad (2024) reported large statistical gains in these areas (n = 130, $R^2 = 0.783$) following a four-week ChatGPT-based module, though context specificity and survey reliance limit generalizability. Similarly, Jamshed and colleagues' (2024) mobile ChatGPT intervention (n = 132) resulted in significant reductions in common grammatical errors, while Polakova and Ivenz (2024) observed improvements (n = 110) in conciseness and passive-voice usage. Mohammed and Khalid's large cohort (n = 322) likewise reported enhanced grammatical accuracy and increased motivation. Furthermore, Yan (2024) noted the importance of teacher scaffolding in achieving the most pronounced writing gains. Regarding overall writing proficiency, the results were mixed. While some studies reported significant improvement through AI integration (Asadi et al., 2025; Bacon & Maneerutt, 2024; Wale & Kassahun, 2024), Alsofyani and Barzanji (2024) did not find statistically significant differences between overall writing performance through AI-generated and teacher feedback. This mixed evidence across the studies underscores the need for caution in over-relying on AI for substantial writing improvements.

To maximize the positive impact of AI-generated feedback, teachers need to support learners' critical engagement with it (Teng, 2025). Indeed, Yan's (2024) report of writing improvement was tied to teacher-supported processing of AI feedback. It is important to note that higher metacognitive awareness correlated with more effective use of AI feedback (Teng, 2025). Furthermore, studies on AI-supported peer feedback indicate that integrating AI tools into the peer review process enhances the quality of feedback provided by students through justification, problem identification, and constructive suggestions, which leads to improvements in the reviewers' own writing skills (Guo et al., 2024; Bacon & Maneerutt, 2024). Yan's (2024) findings (n = 117) show that teacher support is vital for processing AI feedback and sustaining improvements beyond immediate corrections. There were a few successful hybrid protocols where AI and teacher feedback operated in tandem (Asadi et al., 2024; Ghafouri et al., 2024). These insights suggest that effective integration of generative AI requires explicit instruction in prompt formulation, critical evaluation skills, and thoughtfully designed hybrid feedback modes. In other words, optimal writing performance improvements can be attained through the integration of AIgenerated feedback with traditional instruction (Tseng & Lin, 2024; Asadi et al., 2025).

Students' and Teachers' Perceptions of AI-Generated Feedback

Students appreciate the immediacy (Hwang et al., 2024; Abduljawad, 2024; Wale & Kassahun, 2024; Mohammed & Khalid, 2025) and thoroughness of the feedback generated by AI (Abduljawad, 2024; Allen & Mizumoto, 2024; Alsofyani & Barzanji, 2024; Polakova & Ivenz, 2024; Werdiningsih et al., 2024; Mahapatra, 2024). Although these findings are valuable, most

63

studies rely on Likert-scale surveys (e.g., Allen & Mizumoto, 2024; Jamshed et al., 2024), with only a few incorporating in-depth qualitative methods such as focus-group or semi-structured interviews (Kurt & Kurt, 2024; Werdiningsih et al., 2024). Future research should include richer qualitative data and classroom observations to capture the nuanced dynamics of student–AI–teacher interactions.

Despite its usefulness, many students trust teacher feedback more because of its content accuracy and contextual relevance (Zeevy-Soloyev, 2024; Zou et al., 2025). Alsofyani and Barzanji (2024) and Zou et al. (2025) noted that students often prefer teacher feedback for nuanced revisions that reflect their unique writing style. A blended approach combining teacher and AI feedback is preferred so that immediate surface-level errors can be addressed quickly while also not neglecting deeper revisions (Werdiningsih et al., 2024; Asadi et al., 2025). Some studies reported student concerns about over-reliance on AI feedback (Abduljawad, 2024; Werdiningsih et al., 2024; Mahapatra, 2024). Because of its impersonal tone, students were wary of losing their individual writing voice and creativity. Moreover, it was noted that AI feedback introduced unnecessary details or altered intended meanings at times (Alsofyani & Barzanji, 2024). These findings suggest that optimal classroom practice involves strategically blending AI feedback with human review, using AI for quick mechanical fixes while preserving teacher-led, higher-order instruction.

Teachers valued the efficiency of AI-generated feedback to supplement their own. Ghafouri et al. (2024) and Asadi et al. (2025) reported that the promptness of AI feedback can alleviate teachers' workload. However, it was not perceived to replace the personalized feedback provided by experienced educators. Echoing the students' concerns, teachers pointed out the need for careful review before its adoption to avoid inaccurate input. Kurt and Kurt (2024), Yao et al. (2025), and Ghafouri et al. (2024) indicated that additional teacher input is necessary to bridge the gap in a nuanced understanding of student work and provide tailored, context-sensitive guidance.

Hwang et al. (2024) noted that most students use broad and generic prompts when they seek AI-generated feedback, thereby missing the full potential of relevant and constructive reviews. They also observed a misalignment between learners' writing objectives and the prompts they used, indicating ineffective prompting behavior and limited prompt literacy. Furthermore, prompt specificity directly influenced the quality of AI-generated feedback and subsequent revisions. This finding highlights an important instructional gap: the need for systematic training in prompt construction and critical interpretation of AI feedback. Such training could enhance the effectiveness of AI-generated feedback. Therefore, teaching prompt literacy and tasks fostering metacognitive awareness (Teng, 2025) should be integral to EFL writing instruction if AI-feedback is to be effectively integrated into the revision process.

Challenges with AI-Generated Feedback

There is evidence that students with lower metacognitive awareness are more likely to adopt AIgenerated feedback passively (Teng, 2025), leading to over-reliance without critical evaluation. Asadi et al. (2025) reported concerns regarding student dependency on AI feedback, which may hinder the development of independent revision skills. Another challenge is the risk of overwhelming the students with too much AI-generated feedback. Guo and Wang (2024) reported that the large quantity of feedback produced by ChatGPT sometimes included off-topic or redundant comments, which may result in student confusion. Similarly, Zou et al. (2025) noted student preference for teacher intervention in order to receive actionable and specific feedback. As discussed in the previous section, both teachers and students remain skeptical about the ability of AI to capture the nuanced context and writing goals during feedback generation, implying a shortcoming in personalizing writing instruction. Furthermore, improvements in writing may largely be limited to surface-level aspects, rather than in areas such as cohesion and content development involving complex ideas. Taken together, the reviewed studies highlight critical limitations that must be addressed. Specifically, the tendency for AI-generated feedback to overwhelm students and produce off-target suggestions underscores a key practical implication for teachers: AI feedback must be mediated by instructors to prevent student confusion and ensure effective revisions.

CONCLUSION

This systematic review illustrates the transformative capacity of generative AI tools in improving English writing education for language learners. The analysis of 22 studies reveals that AI-generated feedback can significantly improve surface-level aspects of writing, such as grammar, vocabulary, and organization, through immediate feedback that aligns with learning objectives. Nonetheless, the assessment highlights several significant concerns. Although AI-generated feedback effectively resolves mechanical errors and structural issues, it frequently lacks the capacity to provide nuanced, context-sensitive critiques vital for advanced content development in academic writing. Students and teachers have voiced concerns regarding excessive reliance on AI-generated feedback, underscoring the ongoing necessity for human supervision.

The review underscores the importance of incorporating AI-generated feedback into current classroom practices. Although tools like ChatGPT provide immediate and comprehensive corrections for grammar and structure, the involvement of teachers is crucial for tackling more profound content and contextual subtleties (Zeevy-Soloyev, 2024). Integrating AI feedback with teacher and peer evaluations significantly enhances the revision process. Educators should design assignments that require student interaction with AI suggestions, incorporating activities such as prompt formulation and reflective analysis to foster critical evaluation skills and metacognitive awareness. Furthermore, the integration of AI-supported peer-assisted learning, as discussed by Bacon and Maneerutt (2024), enhances collaborative learning environments and helps students cultivate independent revision strategies.

Future studies could investigate the long-term effects of mixed feedback models on writing proficiency and the ways in which teacher scaffolding influences the acceptance of AI suggestions. The variability in educational settings and feedback mechanisms identified in the initial coding calls for comparative analyses of different integration methodologies. Research aimed at cultivating prompt literacy and fostering metacognitive awareness will be crucial, as these competencies seem to significantly facilitate students' ability to critically evaluate and proficiently utilize AI-generated feedback. Moreover, ethical considerations and matters of academic integrity warrant additional examination to guarantee that AI tools enhance rather than detract from the creative and personalized elements of writing.

References

- Abbas, N., Ali, I., Manzoor, R., Hussain, T., & Hussain, M. (2023). Role of artificial intelligence tools in enhancing students' educational performance at higher levels. *Journal of Artificial Intelligence Machine Learning and Neural Network*, 35, 36–49. <u>https://doi.org/10.55529/jaimlnn.35.36.49</u>
- Abduljawad, S. A. (2024). Investigating the impact of ChatGPT as an AI tool on ESL writing: Prospects and challenges in Saudi Arabian higher education. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1–19. <u>https://doi.org/10.4018/IJCALLT.367276</u>

Allen, T. J., & Mizumoto, A. (2024). ChatGPT over my friends: Japanese English-as-aforeign-language learners' preferences for editing and proofreading strategies. *RELC Journal*, 0(0). <u>https://doi.org/10.1177/00336882241262533</u>

- Alsofyani, A. H., & Barzanji, A. M. (2024). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research*, 63(2), 431–463. https://doi.org/10.1177/07356331241307297
- Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity*, 56, 101766. https://doi.org/10.1016/j.tsc.2025.101766
- Bacon, E. D., & Maneerutt, G. (2024). Enhancing English as a foreign language academic writing through AI and peer-assisted learning. *Journal of Institutional Research South East Asia*, 22(3), 282–314. Retrieved from

http://www.seaairweb.info/journal/articles/JIRSEA_v22_n03/JIRSEA_v22_n03_Article14.pdf

- Chan, C., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(43), 1–18. <u>https://doi.org/10.1186/s41239-023-00411-8</u>
- Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*, 00(0), 1– 23. <u>https://doi.org/10.1177/13621688241239764</u>
- Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *Internet and Higher Education, 63*, 100962. <u>https://doi.org/10.1016/j.iheduc.2024.100962</u>
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29(7), 8435–8463. <u>https://doi.org/10.1007/s10639-023-12146-0</u>l
- Hwang, M., Jeens, R., & Lee, H. (2024). Exploring learner prompting behavior and its effect on ChatGPT-assisted English writing revision. *Asia-Pacific Education Researcher*, 1–11. <u>https://doi.org/10.1007/s40299-024-00930-6</u>
 Jamshed, M., Ahmed, A. S. M. M., Sarfaraj, M., & Warda, W. U. (2024). The impact of ChatGPT on English language learners' writing skills: An assessment of AI feedback on mobile. *International Journal of Interactive Mobile Technologies*, 18(19), 18–36. <u>https://doi.org/10.3991/ijim.v18i19.50361</u>

- Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. Journal of Information Technology Education-Research, 23, 1–17. <u>https://doi.org/10.28945/5370</u>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(9), 1–18. <u>https://doi.org/10.1186/s40561-024-00295-9</u>
- Maphoto, K., Sevnarayan, K., Mohale, N., Suliman, Z., Ntsopi, T., & Mokoena, D. (2024). Advancing students' academic excellence in distance education: Exploring the potential of generative AI integration to improve academic writing skills. *Open Praxis*, 16(2), 142– 159. <u>https://doi.org/10.55982/openpraxis.16.2.649</u>
- Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: Mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, 15(7), 1–26. https://doi.org/10.1186/s40468-025-00343-2
- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1), 2410101. <u>https://doi.org/10.1080/2331186X.2024.2410101</u>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British Medical Journal*, 372(71), 1–9. https://doi.org/10.1136/bmj.n71
- Rahman, N., Zulkornain, L., & Hamzah, N. (2022). Exploring artificial intelligence using automated writing evaluation for writing skills. *Environment-Behaviour Proceedings Journal*, 7(SI9), 547–553. <u>https://doi.org/10.21834/ebpj.v7isi9.4304</u>
- Simms, R. (2024). Work with ChatGPT, not against. *Nurse Educator*, 49(3), 158–161. <u>https://doi.org/10.1097/nne.00000000001634</u>
- Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3), 36–57. <u>https://doi.org/10.58304/ijts.20240304</u>
- Teng, M. F. (2025). Metacognitive awareness and EFL learners' perceptions and experiences in utilising ChatGPT for writing feedback. *European Journal of Education*, 60(e12811), 1– 17. <u>https://doi.org/10.1111/ejed.12811</u>
- Tseng, Y., & Lin, Y. (2024). Enhancing English as a foreign language (EFL) learners' writing with ChatGPT: A university-level course design. *Electronic Journal of E-Learning*, 22(2), 78–97. <u>https://doi.org/10.34190/ejel.21.5.3329</u>
- Wale, B. D., & Kassahun, Y. F. (2024). The transformative power of AI writing technologies: Enhancing EFL writing instruction through the integrative use of Writerly and Google Docs. *Human Behavior and Emerging Technologies*, 221377, 1–15. https://doi.org/10.1155/2024/9221377
 - Wang, C. (2024). Exploring students' generative AI-assisted writing processes:
 Perceptions and experiences from native and nonnative English speakers. *Technology Knowledge and Learning*, 1-22. <u>https://doi.org/10.1007/s10758-024-09744-3</u>
- Werdiningsih, I., Marzuki, & Rusdin, D. (2024). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts and Humanities*, 11(1), 2392388. <u>https://doi.org/10.1080/23311983.2024.2392388</u>

- Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology*, 28(1), 1–19.
- Yao, Y., Zhu, X., Xiao, L., & Lu, Q. (2025). Secondary school English teachers' application of artificial intelligence-guided chatbot in the provision of feedback on student writing: An activity theory perspective. *Journal of Second Language Writing*, 67, 101179. https://doi.org/10.1016/j.jslw.2025.101179
- Yeralan, S., & Lee, L. (2023). Generative AI: Challenges to higher education. Sustainable Engineering and Innovation, 5(2), 107–116. <u>https://doi.org/10.37868/sei.v5i2.id196</u>
- Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching and Learning*, 6(3), 1–23. <u>https://doi.org/10.29140/tltl.v6n3.1482</u>
- Zou, S., Guo, K., Wang, J., & Liu, Y. (2024). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning*, 1–30. <u>https://doi.org/10.1080/09588221.2024.2447279</u>

Inanc Karagoz serves as a faculty member at Bartin University, Turkiye, in the Department of English Language Teaching. Her research focuses on reading engagement, peer support, and teacher education.

Email: inanckaragoz@gmail.com